Validating the Validation: The Influence of Liquid Water Distribution in Clouds on the Intercomparison of Satellite and Surface Observations

N. A. J. SCHUTGENS AND R. A. ROEBELING
Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands

(Manuscript received 15 September 2008, in final form 10 February 2009)

ABSTRACT

The intercomparison of LWP retrievals from observations by a geostationary satellite imager [Spinning Enhanced Visible and Infrared Imager (SEVIRI) on board Meteosat Second Generation (MSG)] and a ground-based microwave (MW) radiometer is examined in the context of the inhomogeneity of overcast cloudy skies. Although the influence of cloud inhomogeneity on satellite observations has received much attention, relatively little is known about its impact on validation studies. Given SEVIRI's large field of view (3 km x 6 km for northern Europe), especially when compared to the narrow width of the radiometer tracks (100–200 m), cloud inhomogeneity may be expected to significantly affect the satellite retrieval validation.

This paper quantifies the various validation uncertainties resulting from cloud inhomogeneities and proposes an approach to minimize these uncertainties. The study is performed by simulating both satellite and ground-based observations through resampling a set of high-resolution (100 m) cloud fields that are derived from 1 km x 1 km Moderate Resolution Imaging Spectroradiometer (MODIS) observations. The authors’ technique for generating realistic high-resolution LWP fields preserves the information present in the original observations and creates extra LWP variation at smaller-length scales by considering clouds as simple fractals. The authors believe that this is a new technique for creating high-resolution LWP fields.

Validation errors resulting from cloud inhomogeneity can be classified in two groups. The first group relates entirely to the retrieval process for satellite observations and includes the well-known plane-parallel bias as well as field-of-view mismatches between different channels used in the retrieval. The second group relates to differences in the scene observed by satellite and ground-based sensors. This includes systematic shifts in the observed scene resulting from viewing conditions (parallax effect), offsets between satellite images and ground sites, and different fields of view.

Results indicate that the plane-parallel bias for the authors’ sample of 604 clouds has a median value of 3.3 kg m⁻². All other error contributions appear to be random and have no biases. For individual observations, the parallax effect easily dominates the total error budget for sites that are observed under large viewing angles (e.g., northern Europe). The authors show that this error may be partly compensated by using information about cloud-top heights and by spatially interpolating among an array of SEVIRI pixels to obtain the best estimate of the satellite-retrieved LWP value over the ground site. Optimal intercomparison of satellite and ground-based observations is also possible by matching the track length of the ground observations to the imager’s pixel size in the wind direction.

Thus, one surprising conclusion is that the LWP errors resulting from the second group (scene differences) are significantly larger than those resulting from the first group (satellite retrieval), even after corrections have been applied. Smaller satellite pixels do not alleviate the problem but rather aggravate it, unless the parallax error is corrected. Temporal or spatial averages of observations may be used to reduce the random errors, but the statistical properties of such aggregates are, at the moment, not obvious for reasons that will be discussed. Calibration errors are not considered in the present study.
1. Introduction

Clouds are considered an important but poorly understood aspect of the climate system (Kristjánsson et al. 2000; Potter and Cess 2004). Their presence modulates the radiative balance (Slingo 1990) and therefore the energy available to drive atmospheric dynamics, the hydrological cycle, and global warming. Because cloud radiative properties on a macroscale (cloud albedo and cloud lifetime) are expected to depend on cloud microphysics (Twomey 1977; Albrecht 1989), detailed observations of clouds are a prerequisite for a better understanding of the climate system.

Polar-orbiting satellites allow global coverage on a timescale of a few days, whereas geostationary satellites allow coverage of a significant part of the earth on time scales of tens of minutes. Satellite observations are thus very interesting from the point of view of cloud studies. Minnis et al. (1992) combined surface observations with Geostationary Operational Environmental Satellite (GOES) visual and thermal infrared radiances to determine cloud optical thickness (COT) and cloud-top height. Nakajima and King (1990) were the first to develop an algorithm to derive COT and effective particle size from dual satellite radiance observations (at 0.75 and 2.16 μm). This technique was later applied to Advanced Very High Resolution Radiometer (AVHRR) observations (Nakajima and Nakajima 1995; Kuji et al. 2000). The dual-view capacity of ATSR-2 was employed by Evans and Haigh (1995) to retrieve COT and effective particle size. Currently, Moderate Resolution Imaging Spectro-radiometer (MODIS) flown on the Aqua and Terra platforms is providing detailed information on cloud microphysics (King et al. 1997; Platnick et al. 2003).

The inhomogeneous nature of water cloud macro- and microphysical horizontal structures has received much attention. For a wide variety of cloud types, satellite imagery shows that cloud properties (perimeter, area, and radiance) exhibit a power law in their spectral distribution (Cahalan and Joseph 1989; Cahalan and Snider 1989; Barker and Davies 1992; Lovejoy et al. 1993). The observed break in this power law at small scales (~300 m) was shown to be due to radiative effects and is not an intrinsic property (Davis et al. 1997; Oreopoulos et al. 2000). The same power law has been observed in situ liquid water content (LWC) observations (Davis et al. 1999; Gerber et al. 2001; Siebert et al. 2006) and ground-based radiometer observations (e.g., Feijt and Jonker 2000). Therefore, the radiance variability is commonly understood as resulting from LWC being passively advected in a turbulent medium.

The previously mentioned retrieval techniques all assume homogeneous clouds, so it is no surprise that a lot of research has been devoted to this obviously flawed assumption. When discussing the effect of cloud inhomogeneity on cloud retrievals, it is useful to conceptually distinguish between the plane-parallel bias and the shadowing–illumination error. The plane-parallel bias results from spatial variation of COT [or liquid water path (LWP)] within a pixel and the nonlinear relationship between radiance and COT. The plane-parallel bias is usually dominant at larger-length scales (several kilometers and above). The shadowing–illumination error results from neglecting variations in cloud heights and is typically dominant at smaller-length scales (1 km and below). Cahalan et al. (1994) and Barker and Davies (1992) found that, because of the plane-parallel bias, inhomogeneous clouds have reduced albedo and radiative fluxes when compared to homogeneous clouds with the same average LWP. On the other hand, Iwabuchi and Hayasaka (2002) showed that at small scales shadowing–illumination can cause both under- or overestimation, depending on the scattering geometry (see also Loeb et al. 1998). Using detailed 3D radiative transfer calculations for high-resolution cloud fields, Marshak et al. (2006), Zimmer and Mayer (2006), and Kato et al. (2006) found that, in general, COT was underestimated, whereas the effective size was overestimated when both effects were considered.

For observational studies of the impact on cloud inhomogeneity, see, for example, Várnai and Marshak (2002), who studied shadowing–illumination in the case of MODIS observations, or Dim et al. (2007), who compared observations from a geostationary satellite [Geostationary Meteorological Satellite-5 (GMS-5) Stretched Visible and Infrared Spin Scan Radiometer (SVISSR)] to those from a polar-orbiting satellite (Terra MODIS). Both groups of authors concluded that LWP can be either under- or overestimated, depending on whether one looks at an illuminated or shadowy cloud side.

For theoretical studies of the radiative effects of cloud variability, different cloud models have been employed. Early studies used (bounded) cascade models to generate horizontally variable LWP fields (Scherertz and Lovejoy 1987; Cahalan 1994; Marshak et al. 1994). Later studies were conducted for large eddy simulation (LES) clouds (Marshak et al. 2006; Kato et al. 2006). Zimmer and Mayer (2006) retrieved high-resolution cloud structures from high-resolution cloud observations. Techniques for artificially creating cloud fields based on observations have also been developed (Venema et al. 2006; see Hogan and Kew 2005 for creating cirrus clouds).

Validation of satellite observations with ground observations has been performed by various authors (see, e.g., Nakajima and Nakajima 1995; Kuji et al. 2000; Jolivet and Feijt 2005). But the effect of inhomogeneity on such a validation effort has received little attention.
In this paper, we will study the effect of horizontal LWP and particle-size inhomogeneity on the intercomparison of satellite with ground-based observations. In particular, we will concern ourselves with LWP observations by the Spinning Enhanced Visible and Infrared Imager (SEVIRI) imager on board the geo-stationary Meteosat Second Generation (MSG) satellites and ground-based microwave (MW) radiometers in northern Europe. The analysis is performed by creating realistic high-resolution (100 m) LWP fields (derived from MODIS observations) and then simulating satellite imager and ground-based MW radiometer LWP retrievals. This study is therefore theoretical and allows definition of a truth LWP. Consequently, we can analyze the different error contributions to the validation. One group of errors is entirely due to the assumption of homogeneous clouds in the satellite retrieval. A second group of errors is due to the actual comparison of data from sensors with different fields of view (FOVs).

Water clouds also exhibit vertical inhomogeneity, often displaying (quasi) adiabatic profiles (Brenguier et al. 2000; Pawlowska et al. 2000; Schüller et al. 2003). Clearly, this will impact comparison of LWP measurements by two different sensors, especially as one (satellite) sees the top of the cloud and the other (ground site) sees the bottom. However, the focus of this paper is on horizontal variability, and vertical inhomogeneity (of effective particle size, in particular) will be ignored.

In section 3, we describe the selection of MODIS cloud fields (with a resolution of 1 km) that will be used in section 4 to create realistic high-resolution (100 m) LWP fields. In section 5, we describe how we simulate the LWP validation and consider the various error contributions in detail. This section also considers possible improvements in the validation procedure. Finally, section 6 contains a summary with conclusions.

In this paper, we will discuss error statistics using quantiles, because all error distributions we find are strongly non-Gaussian and some (e.g., the LWP retrieval error resulting from the homogeneity assumption) are strongly skewed. If \( Q_p \) represents the \( p \)th quantile of a distribution, then \( Q_{50} \) is also known as the median. The spread of a distribution is assessed through \( \Delta Q_{68} = Q_{84.2} - Q_{15.8} \) and \( \Delta Q_{95} = Q_{97.75} - Q_{2.25} \), which agree with 2 and 4 times the standard deviation, respectively, in the case of a Gaussian distribution. The word “variability” is taken to mean \( \Delta Q_{68} / Q_{50} \), a measure of the spread divided by the median of the distribution.

2. SEVIRI sensor

MSG is a new series of European geostationary satellites that is operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). In August 2002, the first MSG satellite (Meteosat-8) was launched successfully; in December 2005, the second MSG satellite (Meteosat-9) was launched. MSGs are spinning stabilized satellites that are positioned at an altitude of about 36 000 km above the equator, near the prime meridian (at 9.25°E for Meteosat-8 and 0.16°E for Meteosat-9 in December 2008).

The SEVIRI instrument scans the complete disk of the earth 4 times per hour and operates 12 channels simultaneously. There are three solar channels (0.6, 0.8, and 1.6 \( \mu \)m), eight infrared channels (3.9, 6.2, 7.3, 8.7, 9.7, 10.8, 12.0, and 13.4 \( \mu \)m) and one high-resolution broadband visible channel (0.3–0.7 \( \mu \)m). The nadir spatial resolution of SEVIRI is 1 km \( \times \) 1 km for the high-resolution channel and 3 km \( \times \) 3 km for the other channels. By sensing in narrow and numerous wavelength bands, it is possible to identify specific cloud and surface properties as well as obtain information on the composition and thermodynamic characteristics of the atmosphere. Six of SEVIRI’s channels are similar to those of the AVHRR instrument on board the National Oceanic and Atmospheric Administration (NOAA) and Meteorological Operation (METOP) polar-orbiting satellites.

3. MODIS data

Our purpose in this section is to show how we selected the contiguous MODIS cloud fields that will later be used to create artificial but realistic high-resolution LWP fields. To briefly describe this technique, we propose to use observed LWP variability over a 10 km \( \times \) 10 km contiguous cloud field to represent LWP variability at 100 m after appropriate scaling. At the very least, this requires that we have two datasets: one of cloud fields that serve as the main cloud scene and one of cloud fields that are used to create variability at 100-m scales. In principle, both datasets may contain the same clouds, but the paucity of useful observations over land means that we were forced to construct the second dataset from observations over ocean.

We used the MODIS collection 5 cloud product from both the Terra and Aqua platforms for the period of August 2006 up to and including July 2007 over northern Europe and the North Atlantic. Searching for contiguous cloud fields over land, it became obvious that 25 km \( \times \) 25 km was an optimal field size. Larger (contiguous) fields occur much less in the MODIS dataset, whereas smaller fields simply do not have the spatial coverage required for our analysis (see section 5). All in all, we found 604 contiguous 25 km \( \times \) 25 km cloud fields, which will later serve as scenes.
In addition, we found 1329 contiguous 30 km × 30 km cloud fields over ocean from which we drew the required 10 km × 10 km fields. Searching in the MODIS dataset for contiguous 10 km × 10 km fields (over ocean) does not increase the number of found fields, whereas the 30 km × 30 km fields allow a more robust definition of the internal LWP variability.

The MODIS dataset contains many more (contiguous) cloud fields than we found for this period and location. But we applied other criteria as well to arrive at the most reliable spatial distributions of LWP possible. Those criteria will now be briefly discussed. Relevant information always came from the MODIS MOD06 level 2 cloud product.

First, we looked for 5 km × 5 km fields that were tagged confidently as water clouds (based on infrared observations) and were observed under optimal viewing conditions. We required solar zenith angle \( \theta_s \leq 60^\circ \) and scattering angle \( 100^\circ \leq \Theta \leq 180^\circ \) because Loeb and Coakley (1998) showed that inverting the satellite observed radiances into cloud properties is most reliable for such scattering geometries. Viewing zenith angles were limited to \( \theta \leq 10^\circ \).

Second, we looked at the individual 1 km × 1 km MODIS pixels and required that they were confidently cloudy with no cirrus or heavy aerosol present. To allow for the best possible retrievals, observations over sun glint and/or ice or snow were discarded. Again, cloud phase was checked (the MODIS cloud phase product for 1 and 5 km is based on different algorithms; R. Frey 2007, personal communication).

Finally, the confidence (as assessed by the MODIS team) for the cloud optical thickness, effective particle size, and LWP were all required to be very good. We note that, in the MODIS product, LWP is determined from COT and particle size.

All cloud fields thus found where required to be spatially separated (center to center) by a distance of at least \( \sqrt{2} \) times the field size. This prevented overlap of cloud fields and allowed us to view the cloud fields as independent samples of spatial LWP distributions.

A random subset of the selected 25 km × 25 km clouds over land and 30 km × 30 km clouds over ocean was visually inspected. Some LWP fields over ocean showed a high degree of local clustering that could be indicative of drizzling or even raining clouds in our sample. After demanding that no field has individual pixels with an LWP > 500 g m\(^{-2}\) or \( r_{\text{eff}} > 15 \mu\text{m} \), the number of scenes with such strong clustering was severely reduced.

Looking at the sample averages for the main cloud parameters, we see the following: The observations are equally split among the Terra and Aqua platforms, and they yield similar cloud optical thickness, LWP, and effective particle size. There is a marked land–ocean contrast, which is likely due to sampling, although the difference agrees with that expected for maritime and continental clouds. Median cloud optical thickness (21.7 versus 13.8) and median LWP (128 versus 101 g m\(^{-2}\)) are higher for the clouds over land than over ocean, presumably because our criterion of contiguity implies that only very thick clouds make the grade over land. Median effective particle size is larger over the ocean (11.4 \( \mu\text{m} \)) than over land (8.6 \( \mu\text{m} \)), as has been shown to be the case in many studies. The internal LWP (variability within each field) is somewhat larger for clouds over land than over ocean (0.8 versus 0.6). Overall, the (horizontal) variability in effective particle size (0.2) is much smaller than that of LWP (0.6–0.8).

It is interesting to take a closer look at the internal LWP variability found in the cloud fields over land. From Fig. 1 we see that no cloud field is truly homogeneous, although some show little variation. The distribution is skewed to larger values, implying that strongly inhomogeneous fields will occur more often than just the sample mean and standard deviation would suggest. The intersample \( \Delta Q_{68} \) of the LWP difference as a function of distance within the cloud field is shown in Fig. 2. Note that, if LWP was an independent random variable in each pixel, \( \Delta Q_{68} \) would show a horizontal line. We interpret \( \Delta Q_{68} \) as a measure of correlations in the LWP field and conclude that those correlations are strongest within the first 5 km. Also note the isotropic character of our sample. Finally, from Fig. 3 it is clear that the variability in effective particle size is not nearly as pronounced as the variability in LWP.

We stress that the sample of contiguous cloud fields thus obtained is not a statistically representative selection.
of water clouds. In particular, thin clouds will be underrepresented and broken clouds are, by virtue of their contiguousness, not present. However, we feel that the present sample shows realistic spatial distributions (i.e., not adversely affected by retrieval errors) of LWP. Because this is a comparative study of retrieval errors, this sample should be sufficient.

4. Constructing high-resolution fields

In this section, we will discuss the algorithm used to create high-resolution LWP fields from MODIS observations. High resolution, in our case, means that the new LWP field will have a sampling distance of 100 m (instead of 1 km), but this does not reflect a limitation of the technique.

If we take the common wisdom that clouds are fractal structures to be true, then a simple algorithm for creating high-resolution LWP fields readily presents itself. A fractal, after all, is a "rough or fragmented geometric shape that can be subdivided into parts, each of which is (at least approximately) a reduced-sized copy of the whole" (Mandelbrot 1982). This so-called self-similarity implies that the LWP structure over, for example, 10 km is equally representative of the LWP structure at 1 km or 100 m. As a matter of fact, many simple fractals are created by reproducing a basic structure at smaller and smaller spatial scales (see, e.g., the Koch snowflake or the Sierpinski sieve).

In our case, we start with a single 30 km × 30 km cloud field (observed over ocean) and calculate the interquartile range $\Delta Q_{68}$ of the distribution of LWP difference between MODIS pixels separated by 10 km. This value will be denoted $\Delta_{10}W$ and represents LWP variation at a spatial scale of 10 km. Next, we select the four 10 km × 10 km fields that are located in the corners of this 30 km × 30 km field (there will be little correlation among these fields; see the discussion of Fig. 2). Take one such LWP field $w_{i,j}$ (where $i,j=1,\ldots,10$ are the pixel coordinates) and calculate its mean value $\bar{w}_{i,j}$. A "normalized" distribution of LWP variation can then be defined as

$$G_{i,j} = (w_{i,j} - |w|)/\Delta_{10}W.$$  \hspace{1cm} (1)

After repeating this procedure for all 30 km × 30 km fields, we are left with $1299\times4 = 5316$ fields that can be taken to represent LWP structure at any spatial scale (assuming the fractal character holds at that scale).

Next, we take a 25 km × 25 km cloud field over land (see Fig. 4) and calculate $\Delta Q_{68}$ of the distribution of LWP difference between neighboring MODIS pixels, which we denote $\Delta_{25}W$. We then subdivide each MODIS pixel into 100 m × 100 m subpixels. If the MODIS pixel has an observed LWP of value $w$, then the subpixels can be assigned values

$$w'_{i,j} = w + \Delta_{25}WG_{i,j} \quad i,j=1,\ldots,10,$$  \hspace{1cm} (2)

where $i$ and $j$ are now coordinates of the subpixels relative to the MODIS pixel. Notice that the total LWP of the original pixel remains unchanged because the sum of $G_{i,j}$ is zero. In essence, we use a 10 km × 10 km field ($G_{i,j}$) to represent LWP inhomogeneity at 100 m after
appropriately scaling the field. Note that the original LWP variation at 10 km is scaled to match the variation at 1 km.

What remains to be decided is how to choose a particular field $G$ to fill in a 1 km x 1 km pixel. A first approach would be to select a random field $G_{ij}$ for each 1 km x 1 km pixel. The resulting cloud scene might look like the one shown in Fig. 5. Although the power law is clearly present and extends to shorter spatial scales, the LWP field shows large and arbitrary jumps across the boundaries of the original 1 km x 1 km pixels.

A better approach would be to match the field $G$ across the MODIS pixel boundaries in such a way that the average jump across the MODIS pixel boundary matches the variation among the subpixels. If our database of $G$ fields is large enough, then this should be possible. For instance, one might start by randomly selecting $G$ for staggered 1 km x 1 km pixels, like the black fields on a chessboard. Next, one searches for the appropriate $G$ for the white fields on our chessboard, making sure that the variation across the pixel boundary matches the variation within the pixel. Here, variation is taken as $\Delta Q_{68}$ of distribution of LWP differences between neighboring 100 m x 100 m subpixels (or a similar statistical measure). Experience, however, shows that matching the $G$s in the white chessboard fields to their three or four neighboring fields leaves large residual errors (jumps) unless one has a very large dataset to choose $G$s from.

Instead, we use a different approach, starting with a random $G$ field for the top-left pixel. The $G$ field for the
second pixel from the left on the top line is then matched to this. And then the $G$ field for the third pixel from the left is matched to the LWP field just created and so on. At the end of the top line, move down one line and repeat the process. In this way, $G$ fields only have to match, at most, two neighboring fields, greatly reducing the size of the dataset required. An example of a resulting cloud field can be seen in Fig. 6. This method preserves the power law fairly well. The slight curve present in the graph is very typical of all fields created in the manner described in this paragraph. The chessboard tiling technique, probably because it allows more randomness, shows a straighter line, such as the one in Fig. 5. On a 2.66-GHz Mac Intel, a single 25 km $\times$ 25 km cloud field may be filled in like this in little over 4 min. Much depends on the size of the collection of $G$ fields, which in our case numbered $5316 \times 4 = 21 264$ (four orientations per 10 km $\times$ 10 km field). The implementation was made in interactive data language (IDL).

Creating LWP fields at higher resolution than 100 m is very simple. Either one repeats the present algorithm at smaller scales (e.g., one now fills in the 100 m $\times$ 100 m subpixels, arriving at a final sampling distance of 10 m) or one uses larger size $G$ fields (e.g., the full 30 km $\times$ 30 km, arriving at a final sampling distance of 33.3 m).

We repeat that the final LWP fields are identical to the original MODIS-retrieved LWP fields on a 1-km scale. In addition, the fine structure they exhibit up to scales of 100 m conforms with the oft-observed power law.

5. Modeling the SEVIRI LWP validation

In this section, we discuss how cloud inhomogeneity affects the validation of SEVIRI LWP with ground-based MW radiometer data. We present an approach to model the validation accuracy and discuss the applicability of the approximations inherent in our method. Finally, we present simulated error statistics for the validation of SEVIRI LWP with MW radiometer values.

Deviation between SEVIRI and MW radiometer LWP is due to at least three distinctly different error sources: SEVIRI retrieval, collocation, and FOV mismatch. Retrieval errors for the MW radiometer are ignored; because its measurements are based on thermal emission from the cloud, we expect cloud inhomogeneity to have relatively little effect.

Of these three error sources, only collocation and FOV mismatch errors are due to the validation attempt. Cloud inhomogeneity affects LWP validation because both sensors (satellite and ground-based radiometer) see different parts of the same cloud field at different times. This is a direct consequence of the different spatial and temporal sampling strategies employed by the sensors. As the clouds drift over the observations site, the satellite samples (almost instantaneously) a large two-dimensional view of the cloud from above while the MW radiometer samples a narrow swath from below.

Imagine a cloud field as discussed in the previous section. Such a field represents the LWP distribution at a single moment near a ground site (see Fig. 7). The FOV of SEVIRI is represented by elongated 3 km $\times$ 6 km diamonds (the SEVIRI sensors have a diamond shape; the elongation is due to the SEVIRI view angle of $\theta \sim 60^\circ$, as is the case for Cabauw, Netherlands; Chilbolton, United Kingdom; and Palaiseau, France; see Table 1; Illingworth et al. 2007). The solid diamond in Fig. 7 is the official SEVIRI pixel closest to the ground site. Supposedly, this pixel will compare most favorably with
LWP at the ground site. Averaging LWP over this pixel (taking point spread function into account) would simulate a SEVIRI LWP observation in the absence of retrieval errors. We define an ideal SEVIRI FOV as the imaginary FOV that is centered at a ground site (dotted pixel in Fig. 7). Such an FOV would presumably be best suited for validation purposes.

Because this cloud field is an instantaneous LWP distribution, we need more assumptions to represent the MW radiometer observations. If this cloud field is merely advected without internal evolution (frozen turbulence theorem) for the duration of an observation by the ground site, then the MW radiometer observation may be represented by a narrow track. Note that, for constant integration times, the track will differ with wind speed, wind direction (northwest or southeast in Fig. 7), and cloud altitude. An MW radiometer observation can be simulated by averaging LWP within such a track. Although SEVIRI time movies show it to be a rather poor approximation on time scales of 15 min, the frozen turbulence assumption does allow us to develop a simple framework for our intercomparison of SEVIRI and MW radiometer observations. In actuality, the LWP does not have to be advected passively by the flow. As long as the flow does not modify the statistics of the LWP distribution (so-called stationarity), then our conceptual model should remain valid [see also Feijt and Jonker (2000), who found remarkable agreement in spatial variation of satellite observations and temporal variation in ground observations]. Compared to our current framework, including temporal evolution of the cloud field would only cause an additional random deviation in satellite- and radiometer-derived LWP s. Moreover, this random difference will likely only depend on the radiometer integration time.

Clearly, we will not concern ourselves with the divergence of observations caused by the different viewpoints (from above or below the cloud). This is mainly a retrieval issue and involves detailed studies of radiative transfer. Moreover, it exists independently of cloud inhomogeneity, although the vertical inhomogeneity of effective particle size may be expected to be an important factor.

In this section, we first begin by studying retrieval errors in SEVIRI LWP resulting from inhomogeneity. Second, we study various collocation errors. These two error sources combined allow us to determine an overall SEVIRI error budget for the actual pixel compared to the ideal FOV defined earlier. Third, we consider how (and how much) collocation errors may be reduced by spatial interpolation in a field of SEVIRI pixels. Fourth, we consider errors resulting from the FOV mismatch of the SEVIRI sensor and MW radiometer. Finally, we establish a total error budget for validation. We also present results for a reduced-size SEVIRI pixel to see how improved resolution affects our results.

### a. Retrieval errors

In this subsection, we will look at the effect that inhomogeneity has on the LWP retrievals by SEVIRI. By matching observed radiances at 0.66 and 1.6 μm to those in a lookup table (LUT), COT and effective particle size may be determined; from them, LWP may also be determined. Such LUTs are calculated for homogeneous clouds and SEVIRI is no exception in this respect. In fact, in this paper, we use the same LUTs as used for the SEVIRI operational cloud product (Roebeling et al. 2006).

---

*LWP at the ground site. Averaging LWP over this pixel (taking point spread function into account) would simulate a SEVIRI LWP observation in the absence of retrieval errors. We define an ideal SEVIRI FOV as the imaginary FOV that is centered at a ground site (dotted pixel in Fig. 7). Such an FOV would presumably be best suited for validation purposes.*

![Parallax effect](https://via.placeholder.com/150)

*Fig. 7. A conceptual image of the sampling strategies for SEVIRI and MW radiometer. The view is top-down on a cloud field like the one shown in Fig. 6 (for the sake of clarity, not shown here). The solid diamond is the actual SEVIRI pixel closest to the ground site (dot). Because of the parallax effect, the observed cloud, however, is located to the south (dashed diamond). The dotted diamond centered on the ground site is the ideal SEVIRI FOV. The track running northwest–southeast represents a possible MW radiometer FOV. The inset shows a side view of the atmosphere, explaining the parallax effect.*

### Table 1. Geolocation of the ground sites used in this study.

<table>
<thead>
<tr>
<th>Location (°)</th>
<th>Offset (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lat</td>
<td>Lon</td>
</tr>
<tr>
<td>Cabauw</td>
<td>51.97</td>
</tr>
<tr>
<td>Chilbolton</td>
<td>51.14</td>
</tr>
<tr>
<td>Palaiseau</td>
<td>48.71</td>
</tr>
</tbody>
</table>

Downloaded from http://journals.ametsoc.org/jtech/article-pdf/26/8/1457/3338576/2009jtecha1226_1.pdf by guest on 02 October 2020
Cloud inhomogeneity adds considerable complexity to radiative transfer and the inversion of radiances to cloud physical properties. The errors it causes in retrievals are generally classified as either a plane-parallel bias or a shadowing–illumination error. Typically, shadowing–illumination errors dominate at short spatial scales (∼1 km), whereas the plane-parallel bias dominates at larger scales. It is thus reasonable to assume that the retrieval of LWP for SEVIRI is most affected by the plane-parallel bias.

Under the independent column approximation, the LUTs allow us to calculate radiances at 100-m resolution. After suitably averaging these radiances over the SEVIRI FOV, we obtain simulated SEVIRI observations of radiance at 0.66 and 1.6 μm. Now, the LUTs may again be used, this time to retrieve LWP for the SEVIRI FOV. The discrepancy between actual and retrieved LWP in this FOV is the plane-parallel bias. Note that the independent column approximation is valid as long as the overall error resulting from radiative effects is dominated by this plane-parallel bias. Note also that the lower variability of effective particle size as compared to LWP implies that LWP inhomogeneity is the dominant factor. As a matter of fact, using an $r_{\text{eff}}$ varying over the FOV or an average value causes negligible differences.

It is well known that the plane-parallel bias causes an underestimation of COT, which in turn causes an underestimation of LWP. In this respect, our results are not different (see Fig. 8). The error distribution has a median $Q_{50} = -3.3$ g m$^{-2}$ but, because it is extremely skewed, about 10% of our cloud fields produce individual biases smaller than $-11$ g m$^{-2}$. For our cloud fields, the underestimation of COT is often associated with an overestimation of $r_{\text{eff}}$, somewhat balancing the errors in LWP. Thus, average underestimation of COT is 50% larger than that for LWP, whereas the average overestimation for $r_{\text{eff}}$ is 20% smaller than the LWP underestimation. This overestimation of $r_{\text{eff}}$ is in agreement with previous studies (Marshak et al. 2006; Zinner and Mayer 2006; Kato et al. 2006).

Up to now, we have implicitly assumed that the FOVs for the 0.6- and 1.6-μm channels are identical; however, they are not. Different sensors are used to observe radiances at these wavelengths, and the alignment of the observations is hampered by thermal stresses in the instrument. The difference in the 0.6- and 1.6-μm FOVs can be modeled as a random offset (Gaussian) with a standard deviation of ∼320 m in both east–west (E–W) and north–south (N–S) directions (EUMETSAT 2006, p. 33). For homogeneous clouds, such an offset [the visible–near-infrared (VIS–NIR) wobble] would be harmless, but in the case of inhomogeneity it will lead to additional LWP errors. Its effect can be estimated by simulating the SEVIRI radiances for two different FOVs. The LWP retrieved from these radiances is then compared to the LWP retrieved if the VIS–NIR wobble were absent. These errors are symmetrically distributed around zero and are smaller than the plane-parallel bias (see Fig. 9). Because of the random nature of the VIS–NIR wobble, there is no correlation between this error and the plane-parallel bias. In the next subsection, we will introduce another wobble (pixel wobble) as an error source (Fig. 12, which is discussed later, compares error contributions from both wobbles as a function of wobble size).

As for the effect of scattering geometry on these results, there seems to be none. We have performed the
same analysis for three different scattering geometries ($\theta_0 = 45.7^\circ, 29.8^\circ, 53.2^\circ$; $\phi - \phi_0 = 72^\circ, 4^\circ, 78^\circ$) as might occur for SEVIRI throughout a day (at 900, 1200, and 1600 UTC) but found no differences that might not be explained as statistical noise. The underlying reason for this is that the LUTs themselves do not change drastically with viable viewing geometry.

b. Collocation errors

Collocation errors are due to the different FOVs for the actual observation and for the ideal SEVIRI FOV. They can be further separated into parallax errors and pixel offset/pixel wobble errors.

For ground sites in northern Europe, the $60^\circ$ view angle implies that the observed cloud, located at some altitude above the surface, has a different geolocation than the actual SEVIRI pixel, which is thought to be located at the earth’s surface. Clouds in the Northern Hemisphere are actually to the south of the pixel geolocation to which they are attributed. This parallax effect is explained in the inset of Fig. 7. The arrow represents the view direction of SEVIRI and explains why the FOV is shifted to the south of the pixel. For cloud-top heights of 2 km, this yields a southward shift of 3.46 km, which is on the same order as the SEVIRI pixel size. This so-called parallax error is defined as the difference between averaged LWP in the FOV and in the actual pixel.

Figure 10 shows a histogram of cloud-top heights for Chilbolton for situations in which only water clouds were present. If we randomly assign an observed cloud-top height to each of our cloud fields (assuming constant height over the whole field), then the errors resulting from the parallax effect can be simulated. These errors are symmetrically distributed around zero (even though the southward shift itself has a bias of 2980 m, tan $60^\circ$ times the median of cloud-top height), with $\Delta Q_{68} = 32.3$ g m$^{-2}$.

This result depends strongly on the cloud-top height distribution. In Fig. 11, we examine the influence of a constant shift in cloud-top heights; that is, for a shift of 0 m, the error shown corresponds to the error given in the previous paragraph. For nonzero shifts, we changed the cloud-top distribution of Fig. 10 accordingly and then recalculated errors. So, if clouds are an average of 500 m higher, associated errors will increase by 10 g m$^{-2}$.

The official SEVIRI pixel is usually offset from the imaginary ideal FOV whose center coincides with the ground site (see Fig. 7). The difference between LWP in the official and ideal FOVs (see Table 1) constitutes our estimate of the pixel offset error. In addition, the official pixel does not agree with the actual pixel because of image navigation errors. This so-called pixel wobble error can be estimated by comparing the LWP in the official pixel to that in the actual pixel, because the wobble size is known (EUMETSAT 2006).

The error resulting from the constant offset is estimated to be $\Delta Q_{68} = 21.1, 28.6,$ and 17.2 g m$^{-2}$ for Cabauw, Chilbolton, and Palaiseau, respectively. The error resulting from improper image navigation (we refer to it as pixel wobble) is $\sim 4$ g m$^{-2}$ for all sites.

It is interesting to compare the errors resulting from VIS–NIR wobble and pixel wobble as a function of wobble size. For Fig. 12, we assumed that the pixel wobble (EUMETSAT 2006, p. 32) is similar in north–south and east–west directions at the subsatellite point (SSP) (in actuality, it is not). We see that, for SEVIRI, pixel wobble has a relatively small contribution to the overall error budget. But this error increases rapidly with
increasing wobble size. The error resulting from VIS–NIR wobble is smaller than that resulting from pixel wobble because the NIR channel is more sensitive to the inhomogeneity of effective size (which is small anyway) than of LWP.

**c. Total error budget for SEVIRI**

We are now in a position to assess the total LWP error budget with respect to the ideal FOV resulting from cloud inhomogeneity. Most error sources have distinctly different physical causes and therefore uncorrelated errors. The one exception is the errors resulting from parallax and constant pixel offset. These errors can either amplify or mitigate each other, depending on whether the pixel is located to the north or south of the site. The parallax effect always shifts the FOV southward (at least on the Northern Hemisphere). A summary of the error budget is found in Table 2 and derived from analysis in the previous subsections.

Certain errors, those resulting from LUTs for homogenous clouds or resulting from wobbles, should be identical for each site. The observed discrepancies result from statistical noise. Because pixel offsets are different for each site, we locate each site at a different position within our 25 km × 25 km cloud fields to accommodate all FOVs (official pixel, ideal FOV, observed FOV, and MW radiometer tracks). Consequently, error statistics will vary slightly.

Table 2 shows that the errors resulting from parallax and pixel offset are the dominating contributions. Note, however, that they may partly cancel each other. For Chilbolton, located to the south of the pixel, there is a weak anticorrelation ($r = -0.49$) between the errors resulting from parallax and from site offset. This reduces overall LWP errors somewhat. Cabauw, on the other hand, is located mainly to the west of the pixel and only 155.8 m south of the SEVIRI pixel, and the anticorrelation between parallax and site offset is much less ($r = 0.18$). Finally, for Palaiseau, the offset error is smallest. Even though the geographical distance of the site to the pixel is similar to Cabauw, the relative distance (ratio to the pixel size) is halved. Because Palaiseau has the largest N–S offset, associated errors correlate positively ($r = 0.57$) with the parallax errors. Consequently, overall LWP error is largest for Palaiseau. We point out the extreme non-Gaussian nature of the error statistics, with $\Delta Q_{95}$ being typically 3 times as large as $\Delta Q_{68}$.

Finally, it is interesting to consider the effect a smaller SEVIRI pixel would have on these error budgets. Here, we consider pixel sizes halved in both N–S and E–W directions while at the same time their sampling distances are halved. Also, as part of the improved resolution, we assume that the wobbles have been reduced by 50%. Error budgets are shown in Table 3 and show some surprising results. First, understandably, errors resulting from retrieval and wobble are reduced (note that for a smaller pixel, retrieval errors may have more contribution from shadowing–illumination, an effect we do not model). Errors resulting from parallax are significantly larger as the smaller pixel size offers a smaller chance of cancelling errors. The offset error depends strongly on the exact sampling positions of the pixels (we

| Table 2. Error budget ($g \ m^{-2}$) of SEVIRI-retrieved LWP with respect to the ideal FOV. |
|---|---|---|---|---|---|
| Cabauw | Chilbolton | Palaiseau |
| LUT | $\Delta Q_{68}$ | $\Delta Q_{95}$ | $\Delta Q_{68}$ | $\Delta Q_{95}$ | $\Delta Q_{68}$ | $\Delta Q_{95}$ |
| LUT | 7.8 | 22.2 | 7.2 | 21.9 | 7.1 | 21.2 |
| VIS–NIR wobble | 4.1 | 15.0 | 3.7 | 15.0 | 3.7 | 15.4 |
| Parallax | 32.3 | 115.3 | 35.3 | 111.3 | 31.7 | 111.9 |
| Site offset | 21.1 | 66.5 | 28.6 | 86.0 | 17.2 | 46.5 |
| Pixel wobble | 4.4 | 18.6 | 4.1 | 19.0 | 4.3 | 15.4 |
| Total | 38.6 | 126.7 | 33.6 | 102.4 | 46.1 | 138.0 |

| Table 3. As in Table 2, but for a sensor with improved (2 times) spatial resolution. |
|---|---|---|---|---|---|
| Cabauw | Chilbolton | Palaiseau |
| LUT | $\Delta Q_{68}$ | $\Delta Q_{95}$ | $\Delta Q_{68}$ | $\Delta Q_{95}$ | $\Delta Q_{68}$ | $\Delta Q_{95}$ |
| LUT | 4.8 | 13.2 | 4.6 | 13.3 | 5.3 | 16.2 |
| VIS–NIR wobble | 3.6 | 13.2 | 3.7 | 12.6 | 3.8 | 13.2 |
| Parallax | 49.6 | 177.6 | 52.7 | 170.8 | 52.6 | 181.5 |
| Site offset | 6.7 | 21.0 | 25.3 | 66.8 | 29.0 | 78.6 |
| Pixel wobble | 3.7 | 15.4 | 3.6 | 16.0 | 3.7 | 14.9 |
| Total | 46.8 | 175.1 | 37.6 | 138.3 | 73.6 | 219.5 |
assume the original sampling plus additional pixels halfway). The (anti) correlations between parallax and offset errors increase and total errors are significantly increased.

The preceding analysis should not let one forget that the LWP error budget for SEVIRI observations is partly due to the LUTs for homogenous clouds and the VIS–NIR wobble. We already see here, what will become more obvious in the next section, that validation actually introduces larger errors than the retrieval itself.

d. Spatial interpolation of MSG LWP fields

In this subsection, we investigate how effectively one can perform spatial interpolation on a grid of SEVIRI pixels to alleviate the collocation errors (not the random error resulting from pixel wobble). We ask ourselves the following question: With what accuracy can we obtain the LWP of a virtual pixel from a spatial array of real SEVIRI pixels? This will be useful to know when comparing SEVIRI observations to MW radiometer observations later on, when the virtual pixel will be imagined centered on a ground site.

Spatial interpolation in two dimensions may be accomplished by many different schemes. In practice, we limit ourselves to a few simple schemes, which are as follows: The nearest-neighborhood scheme simply uses LWP at the nearest real SEVIRI pixel. In the bilinear scheme, the LWP values of the four nearest SEVIRI pixels are bilinearly interpolated to obtain the value at the virtual pixel. In the inverse4 scheme, LWP at the four nearest pixels are averaged with weights determined by the inverse of their distance to the virtual pixel. Finally, we also considered a scheme where a Gaussian weighting over the four nearest pixels of the virtual pixel is calculated. This scheme requires a choice for a typical scale length (Gaussian “spread”); experiments show that, for $L = 0.75$, interpolation errors are minimal. From our sample of 604 cloud fields, we may easily determine both the actual LWP in the virtual pixel and the interpolates. As independent variables in the interpolation, we used pixel indices rather than physical distance.

Error statistics $Q_{68}$ are shown in Table 4 for the three ground sites considered in this study: Cabauw, Chilbolton, and Palaiseau. Of the schemes considered, bilinear interpolation and Gaussian averaging seem to be the most accurate, with the bilinear interpolation slightly better (but this may be because of our choice of ground sites or statistical noise). The biases ($Q_{50}$) resulting from these schemes are generally less than 1 g m$^{-2}$.

e. Validation of SEVIRI with MW radiometer observations

Using the frozen turbulence assumption, we can easily model the LWP observations by the MW radiometer and compare these to the SEVIRI observations. If we assume that each MW radiometer observation is based on integration over a fixed time interval, then different wind speeds translate into tracks of various lengths. First, we will compare MW radiometer observations to those for the ideal SEVIRI FOV whose center coincides with a ground site. Error contributions to SEVIRI LWP resulting from retrieval and collocation effects will, for now, be ignored.

In the top panels of Fig. 13, the validation error as a function of track length and wind direction is shown (there are negligible positive biases, which we will not discuss). Two things are readily apparent: errors depend on wind direction and become minimal for an optimal track length. Unsurprisingly, validation for N–S winds at the optimal track length is more successful, because the mean distance of the track to the pixel edges is smaller. The optimal track length depends slightly on wind direction; it is $\sim 7.5$ km for E–W and $\sim 9$ km for N–S, which agrees with the pixel being larger along the N–S direction. For other winds, there is a gradual variation with directions. Away from the optimal track length, validation errors for all wind directions are quite similar.

Overall, errors show a strong non-Gaussian distribution with $Q_{68}$ typically from 2.5 to 3 times larger than $Q_{68}$. Next, we compare MW radiometer LWP for Cabauw to that observed in the actual SEVIRI FOV. This analysis takes into account the errors resulting from retrieval and collocation as well. Errors are now significantly larger, as is shown in the bottom panels of Fig. 13. There is much less difference between the wind directions and track lengths. Errors are distributed in a strongly non-Gaussian fashion $\Delta Q_{68}/Q_{68} = \sim 3$. The smallest expected errors are given in Table 5 and 6.

From our data, we are also able to conclude that there is no significant correlation between retrieval and collocation errors, on the one hand, and FOV mismatch errors, on the other hand, which validates the use of our concept of ideal FOV. As a matter of fact, total error budgets (determined in a separate analysis) are neatly explained by the (squared) sum of their constituting parts (retrieval, collocation, and FOV mismatch errors).

The effect of a reduced SEVIRI pixel is shown in Fig. 14. The top panel shows validation errors for the

| Table 4. Error budget $Q_{68}$ (g m$^{-2}$) for spatial interpolation of MSG LWP fields. |
|-----------------|-----------------|-----------------|
|                | Cabauw          | Chilbolton      | Palaiseau      |
| Nearest         | 22.0            | 24.8            | 17.3           |
| Bilinear        | 9.0             | 13.6            | 11.1           |
| Inverse4        | 21.1            | 18.5            | 14.7           |
| Gauss4 ($L = 0.75$) | 9.3             | 13.2            | 11.9           |
Because of the smaller pixel size, optimal track lengths are shorter and the smallest errors are lower than for the current SEVIRI pixel size (cf. Fig. 13). The bottom panel shows validation errors for the observed pixel. These errors are now larger than for the current SEVIRI pixel because the smaller pixel is more sensitive to LWP inhomogeneity (cf. Fig. 13).

Finally, we consider the improvement in LWP correspondence, when corrections for the parallax error and offset error are introduced. The pixel offset error may be minimized by spatial interpolation in between neighboring SEVIRI pixels, because the location of the ground site with respect to the SEVIRI pixel array is known. The accuracy of various interpolation schemes

<table>
<thead>
<tr>
<th>Pixel location</th>
<th>Cloud top</th>
<th>Pixel</th>
<th>Cabauw</th>
<th>Chilbolton</th>
<th>Palaiseau</th>
</tr>
</thead>
<tbody>
<tr>
<td>No correction</td>
<td>SEVIRI</td>
<td>39.6</td>
<td>38.9</td>
<td>45.2</td>
<td></td>
</tr>
<tr>
<td>Climatological</td>
<td>SEVIRI</td>
<td>26.8</td>
<td>25.2</td>
<td>23.7</td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>SEVIRI</td>
<td>23.8</td>
<td>22.6</td>
<td>19.5</td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>Reduced</td>
<td>19.2</td>
<td>18.7</td>
<td>17.3</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 5. Minimum total error budget $\Delta Q_{\text{tot}}$ (g m$^{-2}$) for comparison of SEVIRI and radiometer LWP for N–S winds. Error for the ideal FOV is 15.4 g m$^{-2}$.**

<table>
<thead>
<tr>
<th>Pixel location</th>
<th>Cloud top</th>
<th>Pixel</th>
<th>Cabauw</th>
<th>Chilbolton</th>
<th>Palaiseau</th>
</tr>
</thead>
<tbody>
<tr>
<td>No correction</td>
<td>SEVIRI</td>
<td>47.1</td>
<td>44.0</td>
<td>58.0</td>
<td></td>
</tr>
<tr>
<td>Climatological</td>
<td>SEVIRI</td>
<td>31.3</td>
<td>34.8</td>
<td>33.8</td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>SEVIRI</td>
<td>29.2</td>
<td>31.4</td>
<td>31.7</td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>Reduced</td>
<td>25.6</td>
<td>30.6</td>
<td>30.1</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 6. Minimal total error budget $\Delta Q_{\text{tot}}$ (g m$^{-2}$) for comparison of SEVIRI and radiometer LWP for E–W winds. Error for the ideal FOV is 30.3 g m$^{-2}$.**
was investigated in the previous subsection. If we know cloud-top height or at least a climatological average, we may similarly correct for the parallax error. Results for the Cabauw station are shown in Fig. 15. In the top panel, it is assumed we have an accurate cloud-top height for each observed cloud. This allows us to reduce minimal LWP errors to 23.8 and 29.2 g m\(^{-2}\) (for N–S and E–W winds, respectively). Note that the latter error is, within statistical noise, similar to the error for the ideal FOV, whereas the first error is significantly larger. In the bottom panel, the median of the cloud-top distribution was used to correct for the parallax error. The \(\Delta Q_{0.68}\) errors are only slightly affected (26.8 and 31.3 g m\(^{-2}\)), but the \(\Delta Q_{0.95}\) errors for E–W wind are significantly higher. Using a climatological average of cloud-top height results in error distributions that are more strongly non-Gaussian and wider. Finally, we see that correcting for pixel offset and parallax in the case of reduced pixels yields smaller errors. Note, however, that the improvement is rather small (~20%).

6. Summary and conclusions

In this paper, we have provided a detailed error budget for the validation of satellite observations of inhomogeneous clouds with ground observations in the case of overcast skies. In particular, we have studied the agreement between LWP retrievals by the SEVIRI imager on board the geostationary MSG satellites and the ground-based MW radiometer in northern Europe. Because both sensors employ different sampling strategies, they effectively see different clouds, although there will be an overlap. If clouds were homogeneous, this would not be an issue, but cloud inhomogeneity strongly affects the validation effort. In this study, we focused on horizontal LWP inhomogeneity of rather
thick unbroken (contiguous) clouds. We do not consider the effects of incorrect calibration on the validation or of vertical inhomogeneity of effective particle size. In addition, we simplify radiative transfer modeling by using the independent column approximation, ignoring the influence of shadowing–illumination on satellite radiances. We argue that this is a valid approximation for the large SEVIRI pixels (3 km \times 3 km at nadir).

Our methodology is based on realistic but artificial two-dimensional high-resolution LWP fields derived from MODIS observations. A new technique for filling in LWP variability at length scales below 1 km was developed that is consistent with current knowledge of cloud variability. The resulting two-dimensional LWP distributions still exhibit the original LWP distribution as observed by MODIS at length scales of 1 km and larger. Some 604 fields of 25 km \times 25 km and a 100-m resolution are used to study the impact of differences in the aforementioned sampling strategies. Consequentially, simulated LWP observations from either satellites or ground sites may be compared directly and with the known truth for a detailed error budget. In particular, we are able to separate the various error contributions inherent to validation: satellite retrieval errors (resulting from the plane-parallel bias and mismatching FOVs between the VIS and NIR channels), collocation errors (resulting from the parallax scene shift and offsets between the satellite pixel and the ground site), and differences in satellite and ground site FOV. Although our selection of contiguous thick clouds may cause underestimation of the errors, relative error sizes should not be affected.

**Fig. 15.** Validation errors in LWP for the Cabauw pixel as a function of MW radiometer track length and wind direction, if parallax and pixel offset are accounted for. (top) Either it was assumed the exact cloud-top height is known or (bottom) a climatological value (median of the distribution in Fig. 10) is used. Errors resulting from statistical noise are indicated with error bars.
We draw several conclusions from this study. First, error distributions are highly non-Gaussian, implying that standard deviations are not the proper statistic to describe LWP errors. In particular, error distributions are wider than the standard deviation would suggest, leading to underestimates of the error if this statistic is used. Instead, we use quantile ranges like \( \Delta Q_{0.8} \) and \( \Delta Q_{0.95} \). Second, all errors that we consider appear to have symmetric distributions around zero (i.e., no bias), with the obvious exception of the plane-parallel bias. The latter error is strongly skewed, always negative, and has a small median value of \(-3.3 \text{ g m}^{-2}\) over our sample of 604 clouds. Furthermore, among the different error sources that we considered, the random errors inherent to the satellite retrieval of LWP over inhomogeneous clouds [e.g., LUTs developed for homogeneous clouds and mismatch of the instantaneous FOV (IFOV) of different satellite channels] are smaller than the random errors because of comparison of satellite and ground-based radiometer derived LWP. In other words, validation of satellite LWP observations on a case-by-case basis introduces larger errors than the retrieval process itself. The total LWP deviation \( \Delta Q_{0.8} \) resulting from inhomogeneity is 40–50 \text{ g m}^{-2} or 30%–40% for our cloud sample. Our study does not include satellite calibration errors (estimated at 20%; Govaerts and Clerici 2004) but, then again, we do not include retrieval errors for the MW radiometer (estimated at 15 \text{ g m}^{-2}; Gaussiat et al. 2007).

Among the random errors resulting from comparison of satellite and ground-based derived LWP, those due to the parallax effect are clearly dominating for a geostationary satellite and ground-stations in northern Europe (latitude \( \sim 50^\circ \text{N} \)). To some extent, these errors may be compensated if cloud-top altitudes are known and a sufficiently extended cloud field exists that allows for spatial interpolation of LWP. But even if the center of the observed cloud coincides with the ground station, different FOVs of the satellite and radiometer allow for larger errors than the retrieval from satellite observations by itself. However, it is possible to minimize those errors by matching the integration time for the radiometer and prevalent wind speeds (and directions) to the satellite pixel size. Not surprisingly, integrated track lengths should resemble the pixel size. For ground sites that are observed under a high viewing angle, distortion of the pixel implies that final error statistics and optimal track lengths are different for east–west winds than for north–south winds.

Surprisingly, validation results do not improve when smaller satellite pixels are used but rather deteriorate further, unless care is taken to correct for the errors resulting from parallax and pixel offset. The reason for this is that larger pixel sizes partly compensate for the parallax error because of overlap of the geolocation of the pixel and its shifted FOV. If, however, parallax and pixel offset errors are corrected for, then smaller satellite pixels allow tighter error budgets. A pixel that is half the size of the SEVIRI pixel (i.e., only 25% of its area) would yield an improvement in LWP validation (decrease in total error budget) of not more than \(-20\%\).

Another way to reduce the random errors would be to average a large number of clouds over time or space. In the first case, one would consider monthly, seasonal, or even yearly averages. In the second case, one would consider all SEVIRI pixels at nearly the same time within a certain radius of the ground site. In both cases, individual observations cannot be interpreted as independent draws from the same constant distribution. Also, we have shown that individual error statistics are highly non-Gaussian, so it is unlikely that random errors will decrease as fast as \( 1/\sqrt{n} \), where \( n \) is the sample size.

Temporal averaging allows a potentially large number of samples to be used but will yield less-detailed information on error statistics; that is, using all available observations allows one to only assess the size of a very general bias but not the effect of, for example, LWP or solar zenith angle on the validation (in our analysis, these parameters do not matter much but for actual observations a correlation between error and solar zenith angle was found). Clearly, there is a trade-off here.

If SEVIRI pixels within a certain radius of the MW radiometer track are used for validation, parallax errors (presumably a constant error distribution over a few tens of kilometers) may be reduced at the cost of increasing the error resulting from FOV differences (whose distribution widens with distance to the track). The total effect on error budget is not immediately clear and warrants further study.

In any case, the error statistics of any validation approach are derived from the underlying individual error distributions presented in this paper.

Acknowledgments. The authors thank Dr. A. Feijt, H. Deneke, and W. Greuell for stimulating discussion. Dr. W. Greuell and two anonymous reviewers provided useful criticism to improve the paper, which we gratefully acknowledge. This study was started while the first author was employed at KNMI and financed through the SYNTHESIS project.

REFERENCES


Roebeling, R., A. Feijt, and P. Stamnes, 2006: Cloud property retrievals for climate monitoring: Implications of differences between Spinning Enhanced Visible and Infrared Imager (SEVIRI) on METEOSAT-8 and Advanced Very High